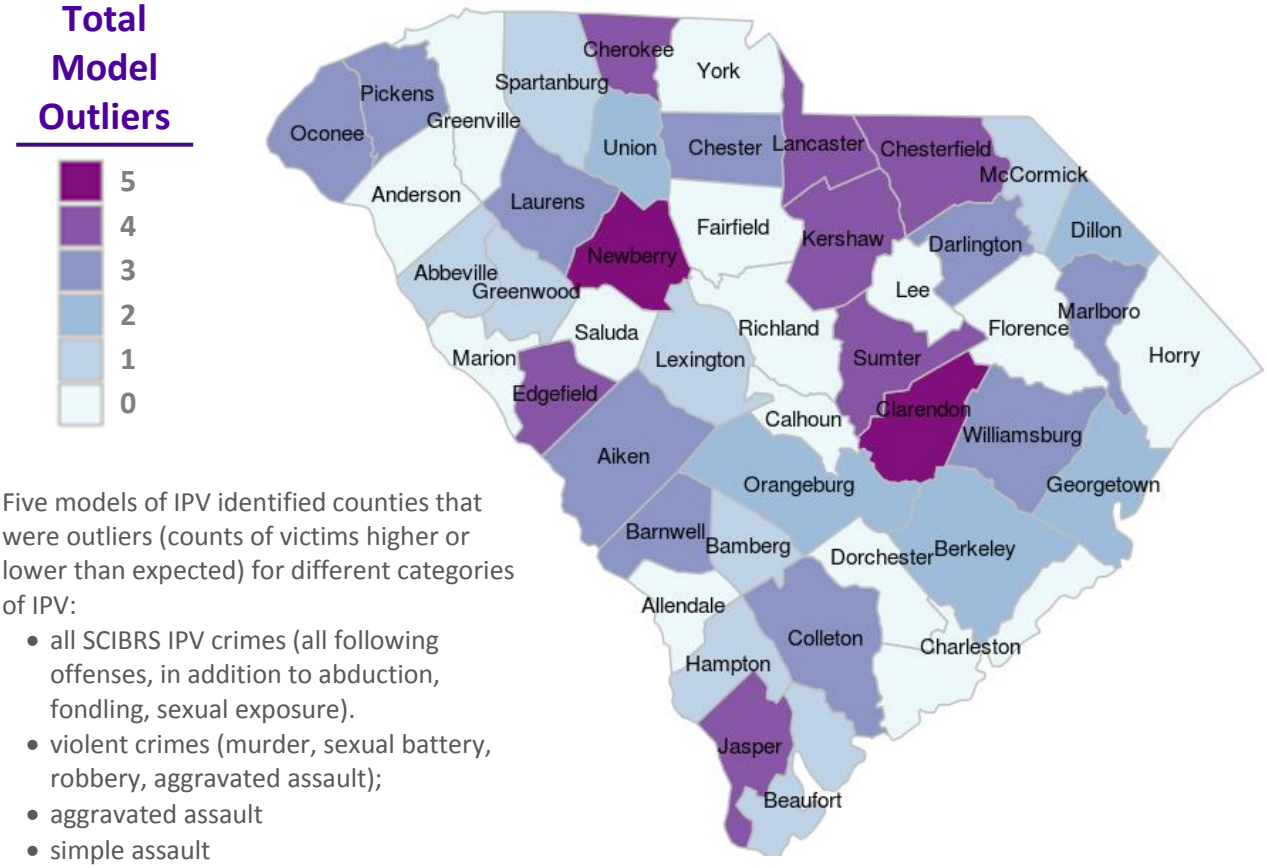
Executive Summary

September 2018

Identifying Intimate-Partner Violence (IPV) County Outliers in the South Carolina Incident-Based Reporting System (SCIBRS) 2011–2015: Sensitivity Analysis by Varying Operational Definitions of IPV

RESULTS AT A GLANCE

SCIBRS 2011–2015 Intimate-Partner Violence



Five models of IPV identified counties that were outliers (counts of victims higher or lower than expected) for different categories of IPV:

- all SCIBRS IPV crimes (all following offenses, in addition to abduction, fondling, sexual exposure).
- violent crimes (murder, sexual battery, robbery, aggravated assault);
- aggravated assault
- simple assault
- intimidation

Source: Stonewall Analytics

The authors are grateful for funding received through the United States Department of Justice (DOJ), Office of Justice Programs (OJP), Bureau of Justice Statistics (BJS) Grant Number 2016-BJ-CX-K022. Any views expressed in this research—including those related to statistical, methodological, technical, or operational issues—are solely those of the authors and do not necessarily reflect the official position or policies of the DOJ, OJP, BJS, or the South Carolina Statistical Analysis Center located in the South Carolina Department of Public Safety's Office of Highway Safety and Justice Programs.

RESULTS IN DEPTH

County-Level Summary of Outliers

Total Model Outliers	County	All SCIBRS IPV Crimes (20 Outliers)	Violent Crimes (11 Outliers)	Aggravated Assault (17 Outliers)	Simple Assault (14 Outliers)	Intimidation (24 Outliers)
1	Abbeville					□
3	Aiken	□		□	-	□
0	Allendale					
0	Anderson	-	-	-	-	
1	Bamberg					□
3	Barnwell			□	□	□
1	Beaufort		-	-	□	
2	Berkeley	-	□	□	-	-
0	Calhoun					
0	Charleston	-	-	-	-	-
4	Cherokee	□	□	□		□
3	Chester	□		□		□
4	Chesterfield	□	□	□		□
5	Clarendon	□	□	□	□	□
3	Colleton	□		□	□	
3	Darlington		□		□	□
2	Dillon	□				□
0	Dorchester	-			-	-
4	Edgefield	□	□	□		□
0	Fairfield					
0	Florence	-			-	-
2	Georgetown	□				□
0	Greenville	-	-	-	-	-
1	Greenwood				□	-
1	Hampton	□				
0	Horry	-	-	-	-	-
4	Jasper	□	□	□		□
4	Kershaw	□		□	□	□
4	Lancaster	□		□	□	□
3	Laurens	□	□		□	-
0	Lee					
1	Lexington	-	-	-	-	□
3	Marion	□		□		□
1	Marlboro	□				
0	McCormick					
5	Newberry	□	□	□	□	□
3	Oconee	□	□			□
2	Orangeburg				□	□
3	Pickens	□			□	□
0	Richland	-	-	-	-	-
0	Saluda					
1	Spartanburg	-	-	-	-	□
4	Sumter		□	□	□	□
2	Union			□	□	
3	Williamsburg	□		□		□
0	York	-	-	-	-	-

Source: Stonewall Analytics

Legend: □ county outlier for that model - county was excluded from analysis for that model

BACKGROUND

South Carolina Law Enforcement Division (SLED) manages the South Carolina Incident-Based Reporting System (SCIBRS), which is National Incident-Based Reporting System (NIBRS)-certified by the Federal Bureau of Investigation (FBI). The SCIBRS stems from approximately 275 law enforcement agencies reporting information about victims, offenses, offenders, and arrestees (if applicable) for all criminal incidents known to police. SLED provides support to agencies through auditing, training, and guidance on coding individual incidents. SLED also stores every incident submitted by the law enforcement agencies on a state repository, submitting those same incidents to the FBI.

The SCIBRS can be used to study a variety of criminal justice subjects. The South Carolina Governor's Domestic Violence Task Force identified it as the primary source for domestic violence data. Accordingly, the integrity of the SCIBRS data must be ensured: quality data best guides policy and the distribution of resources for criminal justice agencies, government institutions, and nonprofit organizations in their mission to aid domestic violence victims. As a NIBRS-certified system, its crimes are categorized by general definitions; thus, the SCIBRS provides a unique opportunity to study domestic violence across jurisdictions—independent of statutory differences.

Because limited resources render it infeasible for SLED to visit all 275 reporting agencies, the South Carolina Statistical Analysis Center (SC SAC) designed a multi-phase research project to aid SLED by statistically guiding its data integrity efforts. This project is the first of its kind in the nation. It develops methodologies that can be shared with uniform crime reporting programs (UCRP) throughout the United States, while also improving lives of domestic violence survivors in South Carolina. It is an attempt to establish a methodology by which UCRP might readily identify counties that are likely to have data quality issues. If successful, this methodology could be used by all such programs to allow for a more focused and efficient assessment of data quality.

In the first phase, Stonewall Analytics (https://www.stonewallanalytics.com/) developed a statistical methodology to identify counties with counts of victims of intimate-partner violent victimization (IPVV) that were "outliers"—either higher or lower than expected. IPVV provides an ideal subset of IPV to develop the modeling methodology because (1) it is the most serious subset of IPV, and so is of deep concern to stakeholders, and (2) the more violent crimes have a better chance of being recorded accurately in police incident data.

During the second and current phase, Stonewall Analytics extends the methodology in the first phase by conducting a sensitivity analysis that uses different subsets of IPV. All definitions focus on intimates (SCIBRS victim-to-offender relationships of spouse, ex-spouse, common-law spouse, (ex-)boy/girlfriend, same-sex relationship), but comprise different categories of SCIBRS offenses that require reporting the victim-to-offender relationship and which occur within the context of domestic violence. Using these five models (with IPV categories listed on the first page of this document) to identify county outliers provides more specificity to the conditions of IPV that render a county an "outlier". Geographic clusters of outliers or single "sore thumb" counties can be mapped and examined in future agency reviews.

METHODS

A random forest model, which is a form of supervised machine learning model, was used to create IPV predictions to be compared against IPV reported values. Random forest models tend to avoid model overfitting; are easy to interpret; and are completed quickly, even on large data sets. Using 2011 national aggravated assault arrest data from the Inter-University Consortium for Political and Social Research (ICPSR), proxy variables for each of the five definitions were used to train, test, and validate the machine learning models. As in the first phase of the project, county-level socioeconomic data from the American Community Survey (ACS) of the United States Census Bureau was incorporated into the modeling. Ultimately, the random forest model predicts which counties are outliers; for this analysis, outliers are defined as counties that have IPV values that are further than one standard deviation from what the model predicts would be the case for the county for at least three separate years ranging from 2011 through 2015. This process was repeated for each of the five definitions of IPV.

A cutoff criterion at the county level was employed to improve the accuracy of the machine learning predictions. While employing a cutoff criterion sometimes excludes counties from the analysis, this cutoff improves the model's overall predictive ability. The criterion excluded any county in the years 2011–2015 where the reported values exceeded the 75th percentile of the proxy variable (see the table on the first page for a listing of excluded counties).

DISCUSSION

This report documents the second phase of a project that assesses the integrity and quality of SCIBRS data at the county level. Performing sensitivity analysis is critical to any robust analysis. There is significant variation in counties deemed outliers based upon the dependent variable selected. For comparison of consistency in the models, only two counties were classified as outliers across all five models. Whereas 11 counties were classified as outliers in the first phase of the project, with the addition of new dependent variables, counties classified as outliers ranged between 11 and 24. When counties classified as outliers were examined geographically, clusters of outliers were also evident. There were small groupings of outlier counties with neighboring borders that demonstrated the same count of outliers. There are several geographic clusters of counties classified with an equal count of outliers. For instance, Lancaster, Chesterfield, Kershaw, and Sumter all exhibit a count of four for outlier status. Aiken and Barnwell were outliers in three of the five models. Additionally, Orangeburg, Berkeley, and Georgetown were outliers in two of five models. In contrast, some counties that are outliers are geographically isolated from counties with similar counts of outliers, e.g., Newberry and Jasper.

In 1987, South Carolina served as the pilot state for the FBI's NIBRS. In 1991, the SCIBRS was the first uniform crime reporting program in the nation to become NIBRS-certified. The outcome of this project gives South Carolina another opportunity to advance uniform crime reporting; its results will be shared with other uniform crime reporting programs. As the FBI moves all states to NIBRS-compatibility in 2021, South Carolina leads the way in data integrity, while also improving the lives of domestic violence survivors.